ABSTRACT

Effectiveness of a recommendation in an Information Retrieval (IR) system is determined by relevancy scores of retrieved results. Term weighting is responsible for computing the relevance scores and consequently differentiating between the terms in a document. However, current term weighting formula like TF-IDF weighs terms only based on term frequency and inverse document frequency irrespective of other important factors. This results in uncertainty in cases when both TF and IDF values are the same for more than one document resulting in same term weight values and hence unable to segregate the terms based on other crucial factors. In this research, we propose a modification of TF-IDF and other term-weighting schemes that weights terms additionally based on the recency of a term, i.e. metric based on the year the term occurred for the first time and the document frequency. We modified the term weighting schemes TF-IDF, BM25 and Universal Sentence Encoder (USE) to additionally consider the recency of a term and evaluated them on three datasets. Our modified TF-IDF outperformed the standard TF-IDF on all three datasets; the improvised USE model outperformed the standard USE on two of the three datasets; the modified BM25 did not perform well against the classic BM25 term-weighting scheme.